Feature Selection using Stepwise ANOVA Discriminant Analysis for Mammogram Mass Classification

B.Surendiran[1], A.Vadivel[2]
Department of Computer Applications, National Institute of Technology, Tiruchirappalli, India
[1]surendiran@gmail.com, [2]vadi@nitt.edu

***Abstract*—In this paper, a feature selection method using stepwise Analysis Of Variance (ANOVA) Discriminant Analysis (DA) is used for classifying mammogram masses. This approach combines the 17 shape and margin properties of the mass regions and classifies the masses as benign or malignant using ANOVA DA. ANOVA DA provides wilk's lambda statistics for each feature and its level of significance. In ANOVA DA the discriminating power of each feature is estimated based on grouping class variable. Principal component analysis (PCA) does feature extraction but it doesn't consider the grouping class variable. The experiment is performed on 300 DDSM database mammogram images. The stepwise ANOVA DA and PCA dimension reduction methods are used to reduce the number of features used. The feature selection using stepwise ANOVA DA is better as it analyses the data according to grouping class variable. Stepwise ANOVA DA based feature selection gives reduced feature set, with high classification accuracy.**

***Keywords*—Discriminant analysis, Digital Mammogram, Shape and margin properties, Classifying Mass as Benign or Malignant, Stepwise ANOVA, PCA**

I. INTRODUCTION

The breast cancer is the leading cause of death in female population. Every 3 minutes, a woman is diagnosed with breast cancer, and in every 13 minutes a woman dies from breast cancer [1]. Mammography is one of the best known technique for early breast cancer detection. Breast cancer death rates have been dropping steadily since 1995 due to earlier detection and increased use of mammography [1]. Computer Aided Detection (CAD) systems have been developed to aid radiologists in diagnosing cancer from digital mammograms and improves breast cancer diagnostic accuracy rate by 14.2% [2].

In breast, malignant and benign are abnormal growth of tumor cells. While malignant are considered as cancerous tumors, the benign are non-cancerous. According to Breast Imaging Reporting and Data System (BIRADS) the masses are characterized by shape, size, margins (borders) and density [3]. Benign masses are round and oval in shape and have smooth, circumscribed margins. The malignant masses have irregular shape and ill-defined, microlobulated or spiculated margins. It has been observed that shape and margin characteristics can be effectively used for classifying the masses either as benign or malignant. Based on shape and margin properties, some of the known approaches which classify the abnormalities based on BI-RADS system have been giving accurate results [4, 5]. Thus, in this paper, mass shape and margin properties are given high importance. These simple and yet effective geometric shape and margin properties visualize the masses as the way radiologists visualize the mammograms.

Researchers had proposed various features for classifying masses in mammograms. The statistical features like uniformity, smoothness, third moments etc which utilize gray value or histogram of masses are used for classifying the masses [6]. However the gray values of mammogram tend to change, due to over-enhancement or in presence of noise. Most of the existing works have been concentrated on classifying the mass as normal or abnormal using shape features [7, 8]. But, most of previous approaches which classify the mass as benign or malignant are not able to get very good classification rate. In [9], a complex Bayesian Neural Networks classifier with 5 statistical measures has been used to classify the masses. The test has been carried out with small dataset containing only 17 sample mammograms and have achieved maximum of 81% accuracy.

In this paper, 17 shape and margin properties are used for classifying the mass either as benign or malignant. It has been observed that not all the properties are equally important. The dimension or number of features can be reduced, which simplifies the classification. Dimension reduction techniques are feature selection and feature extraction. PCA is the commonly used feature extraction method in the literature [10-12]. The main disadvantage of the PCA method is does not consider the grouping class variable. A better feature selection method using stepwise ANOVA discriminant analysis is compared with PCA. The main advantage of ANOVA based feature selection is that, ANOVA estimates wilk's lambda statistic based on the grouping class variable. It performs essential feature selection rather feature extraction without much loss in classification accuracy. The stepwise ANOVA DA is found to be giving good results compared to PCA. This paper is organized as follows. In Section 2, we present feature extraction using shape properties. Next in Section 3, we discuss about ANOVA discriminant analysis classification method. In section 4, we present the experimental results using PCA and stepwise ANOVA DA feature selection method. In section 5, we conclude the paper.

II. MASS SHAPE AND MARGIN FEATURE EXTRACTION

A. Mass Shape Characteristics

According to BIRADS system, the shapes of the masses are characterized as round, oval, lobular, and irregular. Similarly, margin of the masses are characterized as circumscribed, obscured, micro-lobulated, and spiculated margins. Benign masses have round, oval and lobular

DOI: 01.IJSIP.02.01.195

shapes with circumscribed margin. The malignant masses have lobular and irregular shape with ill-defined, microlobulated or spiculated margins. These shape characteristics can be used for classifying the masses present in mammogram.

B. Shape Properties

For the experiments mammograms from DDSM Database [13] are used. And the ground truth available with each mammogram is used to measure the classification rate. Total of 300 (150 benign and 150 malignant) mammograms containing masses is considered for experiment. The masses with various mass characteristics like oval, round, lobular, irregular, architectural distortion, asymmetric density etc are considered. The mass shapes like irregular, nodular, architectural distortion are difficult to measure compared to round, oval masses.

This paper uses 17 shape and margin features extracted from the Region of Interest (ROI) chosen from the mammograms. Details of all 17 features can be found at [14]. These 17 features are Area (A), Perimeter (P), Maximum Radius (Rmax), Minimum Radius (Rmin), Euler Number (EULN), Eccentricity (Ect), Entropy (Entpy), Equivdiameter (Eqd), Elongatedness (En), Circularity (C1 and C2), Compactness (CN), Dispersion (Dp), Thinness ration(TR), Standard deviation of ROI (SD), Edge Std deviation (Esd), and Shape Index(SI).

For each ROI, features are extracted and constructed as feature vector. Shapes feature vector, SFV= {mammogram, $p_{t,}$ mass_type}, where t=1…17 and p_t is the shape property. Features like C1, C2, TR, Eqd, Ect, and CN are used to measure shape characteristics. Similarly for margin characteristics, features such as Entpy, SI, Esd, etc are used. Figure 1 shows the extracted ROI from sample benign and malignant masses.

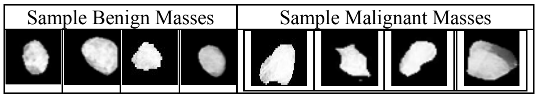

Figure 1. Sample benign and Malignant Masses

C. Shape Property Values

Table I shows the 6 shape feature values out of 17 extracted features for both benign and malignant masses from sample test mammograms.

Table I Various Shape Property Values

Mammogram	**Area**	**Eqd**	**Entpy**	**C1**	**CN**	**SI**
Benign Masses						
Benign1	542	26.27	0.093	0.525	65.21	3.758
Benign2	694	29.726	0.166	0.625	47.729	3.825
Benign3	901	33.87	0.163	0.901	14.172	3.004
Malignant Masses						
Malignant1	2530	56.756	0.38	0.713	106.877	6.529
Malignant2	2727	58.925	0.406	0.723	92.779	6.166
Maligant3	425	23.262	0.089	0.338	237.939	4.613

III. ANOVA DISCRIMINANT ANALYSIS

ANOVA uses single dependent continuous variable, but use more than one independent categorical variable. ANOVA DA is an excellent method as it compares the relation between groups and within-groups. A detail explanation of ANOVA DA can be found at [15].

ANOVA is a special case of the General Linear Model y = Xb + e. Where y is a dependent variable (DV), X is a matrix of predictors or Independent Variables (IVs), b is a vector of regression coefficients (weightings) and e is a vector of error terms. The ANOVA is a procedure that determines whether differences exist between two or more population means by analyzing the within-group and between group variances.

The ANOVA DA classifier predicts the discriminating power of each feature using the wilk's lambda measure. Lower the wilk's lambda value, higher the discriminating power. Wilks lambda is used to test the null hypothesis that the populations have identical means on *D (discriminating function)*. Wilk's lambda is defined as the ratio of within-group sum of squares to the total sum of squares. Wilk's lambda, $\Lambda = \frac{SS_{within\_groups}}{SS_{total}}$.

IV. EXPERIMENTAL RESULTS

The SPSS package is used for applying ANOVA DA, PCA and stepwise ANOVA DA feature selection. The shape property vector with mass_type is given as input for ANOVA classifier. It uses the mass_type (with 0 for benign and 1 for malignant) as dependent variable and all shape and margin features as independent variables. A Leave One Out Cross Validation (LOOCV) technique is used for validation.

The wilk's lambda Λ statistic for each feature is shown in Table II. A significance value p <0.05 shows that the variable has good discriminating power.

Table II Test of Equality of Group Means and PCA components

Features	Λ	p	CDF Co-eff	PCA Components			
Area	.698	.000	.001	0.89	0.16	0.24	0.17
perimeter	.716	.000	-.342	0.92	-0.04	-0.04	0.20
Rmax	.663	.000	-.374	0.89	0.33	-0.07	-0.13
Rmin	.857	.000	.088	0.42	0.53	0.50	-0.27
EULN	.991	.107	-.009	-0.18	0.26	-0.22	-0.77
ECT	.971	.003	.006	0.26	0.58	-0.50	0.31
Eqd	.558	.000	3.003	0.94	0.08	0.24	0.03
En	.968	.002	-.217	-0.25	-0.66	0.66	-0.02
Entpy	.677	.000	-2.153	0.91	0.15	0.25	0.14
C1	.965	.001	-.067	-0.27	-0.67	0.66	0.04
C2	.996	.296	-.196	-0.09	0.55	0.68	-0.11
CN	.914	.000	-.529	-0.62	0.60	0.40	0.15
Dp	.799	.000	.303	-0.66	0.46	-0.12	0.38
TR	.953	.000	.578	-.534	.593	.395	.289
SI	.921	.000	.347	.532	-.707	-.049	.201
SD	.532	.000	.706	.807	.326	.168	-.157
Esd	.616	.000	-.188	.963	-.139	-.004	.040

From the Table II it can be observed that features like Rmax, Eqd, Entpy, SD, Esd, A and P have lower Λ and have high discriminating power compared to other features. The standardized canonical discriminant function coefficients for each features is shown in Table II. The overall wilk's lambda value for the derived canonical discriminant function is 0.380. All the 17 features produce classification accuracy of 86.7% using ANOVA DA classifier.To reduce the number of features, feature extraction method like PCA is used. PCA computes new factors or components from all feature set. From 17 shape and margin features, PCA extracts 4 components is shown in Table II. The extracted four PCA components are classified using ANOVA DA and the classification accuracy achieved is 82%.

V. Feature selection using Stepwise ANOVA DA

In stepwise ANOVA DA feature selection method, a new feature is added to the set and it stops adding features when there is no improvement in overall accuracy. The experimental result of ANOVA DA stepwise method using Wilk's lambda is shown in Table III.

Table III Stepwise ANOVA DA Feature Selection Result

Step	Variables Entered	Wilk's Lambda
1	SD	.532
2	Eqd	.497
3	Entpy	.419
4	Dp1	.404
5	EULN	.398

From 17 feature set, by applying stepwise feature selection a set of 5 features (SD, Eqd, Entpy, Dp, EULN) are selected which gives good classification accuracy of 87.3%. The overall wilk's lambda statistic for selected 5 features is 0.398. Comparison between ANOVA DA, PCA and Stepwise ANOVA DA are shown in Figure 2 and Table IV.

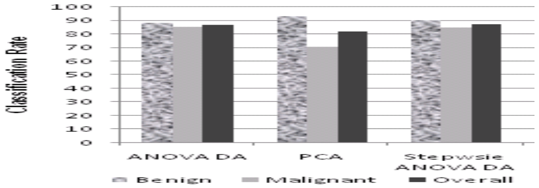

Figure 2. Comparison between various methods

VI. Conclusion

The shape and margin properties of the masses are extracted and used for classification. Vital features are selected using stepwise ANOVA DA. The results are compared with dimension reduction techniques like PCA. The number of features selected or features extracted by PCA and stepwise ANOVA DA are shown in Table IV. The classification accuracy by stepwise ANOVA DA is 87.3% with 5 set of features compared to 82% classification accuracy by PCA with 4 extracted components. Stepwise ANOVA DA method performs better than PCA.As a future work, neural network classifier can be used to test the classification accuracy.

Table IV Comparison of various methods

Method	Variables	Benign	Malignant	Overall
ANOVA DA	17	88%	85.3%	86.7%
PCA	4	93.3%	70.7%	82%
Stepwise ANOVA DA	5	90%	84.7%	87.3%

Acknowledgment

The work done by Dr. A.Vadivel is supported by research grant from the Department of Science and Technology, India, under Grant SR/FTP/ETA-46/07 dated 25th October, 2007.

References

[1] A. C. S. (AMS). "Learn about breast cancer", 2006. http://www.cancer.org.

[2] "Computer-aided Detection Improves Early Breast Cancer Identification". Medical news today. http://www.medicalnewstoday.com/articles/48719.php

[3] "The ACR Breast Imaging Reporting and Data System (BI-RADS)". American College Radiology, 1998. Third Edition, http:// www.imaginis.com/pro/breast_imag_resrc/acr-birads.asp

[4] Markey M. K., Lo J. Y., Tourassi G. D., Floyd C. E.," Cluster analysis of BI-RADS descriptions of biopsy-proven breast lesions", In: Medical Imaging: Image Processing, Proceedings of SPIE Vol. 4684,pp. 363-370 (2002)

[5] Mehul P. Sampat, Alan C., Bovik B., Mia K. Markey," Classification of mammographic lesions into BI-RADS shape categories using the Beamlet Transform", Medical Imaging: Image Processing, Proc. of the SPIE, vol. 5747, pp.16-25, (2005)

[6] Vibha L., Harshavardhan G. M., Pranaw K., Deepa Shenoy P., Venugopal K. R., Patnaik L. M.," Classification of Mammograms Using Decision Trees", In: 10th International Database Engineering and Applications Symposium (IDEAS'06). Pp.263-266 IEEE (2006)

[7] Beatriz A. Flores, Jesus A. Gonzalez," Data Mining with Decision Trees and Neural Networks for Calcification Detection in Mammograms", In: Third Mexican International Conference on Artificial Intelligence, Proceedings -LNCS ,Springer, pp. 232-241,(2004)

[8] Sun Y., Babbs C., Delp E.," Normal Mammogram Classification Based on Regional Analysis", In: Proceedings of the IEEE Midwest Symposium on Circuits and Systems., Vol 45, pp.375-378, (2002).

[9] Leonardo de O. Martins, Alcione M. dos Santos, Arist´ofanes C. Silva1 and Anselmo C. Paiva1, "Classification of Normal, Benign and Malignant Tissues Using Co-occurrence Matrix and Bayesian Neural Network in Mammographic Images", SBRN'06, IEEE computer society, pp. 24-29 (2006)

[10] A.K. Jain, R.P.W. Duin, J. Mao, "Statistical pattern recognition: a review", IEEE Trans. Pattern Anal. Mach. Intel. 22 (1) (2000) 4–37.

[11] Papadopoulos, A., Fotiadis, D.I., Costaridou, L., "Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques" ,Computers in Biology and Medicine,pp:1045 – 1055,(2008)

[12] Zheng B. "Computer-Aided Diagnosis in Mammography Using Content-Based Image Retrieval Approaches: Current Status and Future Perspectives". Algorithms. 2009; 2(2):828-849

[13] Chris Rose, Daniele Turi, Alan Williams, Katy Wolstencroft, Chris J. Taylor," Web Services for the DDSM and Digital Mammography Research", pp. 376-383 (2003)

[14] B.Surendiran, A.Vadivel, Y.Sundaraiah, "Classifying Digital Mammogram Masses Using Univariate ANOVA Discriminant Analysis ", ARTCom 2009,IEEE computer society, Oct 2009. (Best Paper Award)

[15] Aviva Petrie,Caroline Sabin, Book: Medical Statistics at a Glance, 2000

DOI: 01.IJSIP.02.01.195